Taxpayer response to an increased probability of audit: some evidence from Italy

Carlo Fiorio[1] and Alessandro Santoro[2]

February 28, 2012

[1]DEAS, University of Milan, Milan, Italy. email: carlo.fiorio@unimi.it
[2]DSGE, University of Milan-Bicocca, Milan, Italy. email: alessandro.santoro@unimib.it

Abstract

Since 1998 Italy has adopted a method to audit small businesses (Studi di Settore), which defines the probability of tax audits based on presumptive and reported levels of output. In 2007 a letter campaign was implemented by the Italian Tax Agency aimed at reducing manipulation of input reports for tax purpouses threatening that if the "anomaly" was not removed with the 2008 tax declaration, the probability of a thorough tax audit would have drastically increased.

In this paper we analyse a large data set produced by the Tax Agency for this project and made of about 50,000 treated firms and 150,000 controls using Coarsened Exact Matching (CEM) methods to control ex-ante for imbalance.

We find that the letter campaign had a positive and statistically significant effect looking on the average treatment effect on the treated.

PRELIMINARY VERSION, PLEASE DO NOT QUOTE WITHOUT PERMISSION

Keywords: Business Taxation, Tax Compliance, Coarsened Exact Matching, Studi di Settore.

JEL: H26, H25, C13

1 Introduction

Since 1998 Italy has adopted a method to audit small businesses (firms and professionals) known as Studi di settore (Sds). By using this method the probability of audit is increasing in *presumptive* and decreasing in *reported* level of output. Presumptive output is obtained in two steps. First, the Tax Agency (TA) estimates the weighted average productivity of a set of selected inputs within the economic branch of operation of the business, yielding a vector of estimated productivity parameters. Second, the value of inputs is reported by the firm and presumptive output is obtained by multiplication of the vector of productivity parameters by the vector of inputs. Since the vector of productivity parameters is known to the taxpayer at the time he is asked to report inputs, the method is prone to manipulation by taxpayers who can lower presumptive output, and thus audit probability, by underreporting the value of selected inputs.

Relating this method to those known in the literature, we can treat Sds as signals informing the TA about the true level of output but, differently from what is commonly assumed (Macho-Stadler and Pérez-Castrillo 2002), the realization of this signal depends on, and is known to, taxpayers, who can manipulate it ex-ante.

Up until 2005, the method was implemented by the Italian TA without paying any attention to this manipulation bias. As a result, the probability of an Sds-based audit decreased rapidly, and this was interpreted, rather than as a sign of increased compliance, as the direct consequence of the intense activity of underreporting of input values undertaken by a large number of taxpayers.

Since 2005, the TA has reacted to the likely manipulation activity of firms by planning a number of administrative actions. Among these, we consider the initiative known as *Comunicazioni anomalie studi di settore* (Communications on anomalies concerning Studi di settore) which was implemented in tax year 2007. It consisted in sending a letter to taxpayers who allegedly manipulated their report, according to information available at the TA, informing taxpayers that some input data they reported for tax year 2007 were seen as 'anomalous' and that, if not emended for tax year 2008, it would cause the inclusion of the taxpayer in a list of taxpayers to be audited.

We examine here the taxpayers' response to this letter using a large data base of firms' tax reports produced by the TA for this project. We observe data of one third of all treated firms in 2007 (the letter campaign year), and in 2006 and 2008. We also observe a sample of over 150,000 control firms that allows us to apply statistical matching conditioning on observable caracteristics before the campaign was implemented.

Data are analised using the recently developed Coarsened Exact Matching (Iacus, King, and Porro 2011b), which allows us to control the level of imbalance ex-ante, reducing the bias and increasing the efficiency of the estimation of the average treatement effect on the treated.

The paper is organized as follows. Section 2 summarizes the literature which has examined initiatives adopted by Tax Agencies, namely in the US and in Denmark, to increase the perceived probability of being audited by taxpayers. Section 3 describes Sds-based probability of audit and provides definitions of three different statuses: reliability (*coerenza*), consistency (*congruità*) and manipulation. Section 4 derives some theoretical insights by modelling the letter campaign as a change in the probability of audit as perceived by taxpayers. Sections 5, 6 and 7 are devoted to data description, to the matching methodology adopted and to the discussion of empirical results, respectively. Section 8 discusses main results and concludes.

2 The letter campaign and some related literature

The use of letters to increase perceived audit probabilities is not uncommon among Tax Agencies. In particular, letters were used in some field experiments conducted in recent years. Here we shall briefly discuss those documented in Blumenthal, Christian, and Slemrod (2001), which we describe as Minnesota 1, in Slemrod, Blumenthal, and Christian (2001) which we describe as Minnesota 2 and in Kleven, Knudsen, Kreiner, Pedersen, and Saez (2011) which we describe as the Danish Experiment. In the Minnesota 1 experiment a sample of 1700 taxpayers (treated sample) who filed a tax return for year 1993 is randomly extracted from the population of Minnesota taxpayers. The sample is randomly selected using as stratification criteria an income criterion and an opportunity of evasion criterion: income is splitted into high, medium and low, while opportunity of evasion is deemed to be low when the income is subject to third-party reporting and high when there is no such option. Taxpayers included in the treated sample received a letter warning them that their tax returns for year 1994 would be 'closely examined'. Their reporting behaviour is compared to that of a control sample formed by approximately 23000 taxpayers extracted from the stratified population of Minnesota taxpayers who filed a tax return for year 1993. Main results of this experiment are overall quite deceptive. A partially significant positive impact of the letter in terms of average reported incomes (and taxes) for some of the subgroups, namely those with low and average incomes, is offset by a very low impact among taxpayers whose opportunity to evade is low and even a significant *negative* impact of the letter on average reported incomes (and taxes) for the group of high-income taxpayers. Moreover, there is a lack of significance of almost all regression coefficients in both samples.

These results have been interpreted by Blumenthal, Christian, and Slemrod (2001) as follows:

a) for all taxpayers the threat of an audit could have been non credible;

b) the negative impact on high-income taxpayers could be partly explained by the fact that the majority of them have an high opportunity to evade (since no third-party reporting is available for this kind of taxpayers). However, this explanation does not hold for high-income taxpayers who have low opportunity to evade but, despite that, react negatively to the letter.

In the Minnesota 2 experiment two samples (treated samples) each of approximately 20000 taxpayers are randomly selected from the population of Minnesota taxpayers who filed a tax return for year 1993 . The first sample received a letter named as Support Valuable Services whose meaning was that taxpayers should comply voluntarily in order to support the provision of socially valuable activities. The second sample received a letter named as Join the Compliant Majority, whose message was that if one wished to belong to the majority community of citizens one should comply with the tax laws. The reporting behaviour of these two samples is compared to that of a control sample formed by approximately 20000 taxpayers randomly extracted from the population of Minnesota taxpayers who filed a tax return for year 1993. The methodology is very similar to the one adopted in the Minnesota 1 experiment. Again, results do not seem to support the idea that letters are not effective in perceived audit probabilities. Both treated samples report a higher increase in average reported income with respect to the control group, but neither of them are significant. In the regression, dummies denoting groups are insignificant either when evaluated alone or when interacted with other variables. Two somewhat alternative explanations are offered by Slemrod, Blumenthal, and Christian (2001):

a) either the impact of the letters on ethical and social values has been negligible since some expressions used in the letter were ambiguous and could have reinforced the sense of impunity by tax evaders

b)or these values have a modest impact on compliance so that Tax Agencies should not rely upon them to increase taxpayers'loyalty.

Finally, the Danish experiment is accomplished in two steps. In the first one, taxpayers are divided into 2 groups: a first who is audited on their tax returns for tax year 2006 without being previously alerted and a second group who is not audited. In the second part of the experiment, which concerns tax returns for tax year 2007, dependent workers belonging to both groups as previously described are divided in 3 new groups; a first group who receives a letter stating that they will surely be audited (100%-letter); a second group who receives a letter stating that they will be audited with a percentage of 50% (50% letter) and a third group who does not receive any letter. The experiment is complex in its structure and in its objectives. Here we limit the attention to results concerning the impact of the letters on income reported in the second experiment. The main finding of the paper is that such an impact is positive and significant, and, in particular, that it is higher for those dependent workers who were not audited in the first part of the experiment.

3 A description of Italian Sds

Since 1998, Italy has adopted Sds to audit businesses (firms and professionals) conducting an economic activity on a small scale, i.e. reporting an annual output below 7,500,000 euros. Sds can be seen as a method to base the audit probability function on the comparison between presumptive and reported output.[1] To describe it, we first focus on the derivation of presumptive output for each business and then on the characterization of the audit probability function.

As our empirical analysis uses data about firms only, we briefly describe how Sds work for firms (corporated and unincorporated companies and individual entrepreneurs). The TA collects information on structural variables (e.g., size of offices and warehouses, number of employees, main characteristics of customers and providers, etc.) and on accounting variables (mainly referring to amount and cost of inputs and the value of output). A number of statistical analyses are performed to identify and prune the outliers, to group firms in clusters within each business sector, and to select inputs that are statistically more significant in explaining the variance of reported output within each cluster of firms. Then, for each cluster within a business sector, the presumptive productivity of each input is calculated. Presumptive output is finally obtained for every firm as the weighted sum of the reported value of selected inputs, where weights are the presumptive productivity parameters. In turn, these parameters are calculated by the TA on the basis of data reported in previous years (no more than 3) by a subset of firms deemed to be reliable (*coerenti*) in providing relevant information.

More formally, the TA, after dividing business sectors into C clusters and allocating each firm to a single cluster, selects within each cluster $c =$

[1] For a more detailed description and analysis of SdS, see Santoro and Fiorio (2011) and Santoro (2008).

$\{1, 2, ..., C\}$ the group of firms that it believes to be reliable, $R_c \subseteq I_c$, in year t, where I_c is the subgroup of the total population I belonging to cluster c, where $\cup I_c = I$. Hence, it estimates the relation:

$$y_{r,t-3} = \boldsymbol{\beta}'_{t-3} \cdot \mathbf{x}_{r,t-3} + \epsilon_{r,t-3} \tag{1}$$

where $r \in R_c$, $\mathbf{x}_{r,t-3}$ is the $J \times 1$ vector of input, $y_{r,t-3}$ is the value of output reported by firm r, and $\epsilon_{r,t-3}$ is an idiosyncratic error of firm r in period $t-3$, respectively. $\boldsymbol{\beta}_{t-3}$ is the $J \times 1$ vector of unknown productivity parameters, which – once estimated by using standard regression techniques – is denoted $\widehat{\boldsymbol{\beta}}_{t-3}$. Finally, the TA defines the $J \times 1$ vector of productivity parameters coefficient at time t as $\mathbf{b}_t := \widehat{\boldsymbol{\beta}}_{t-3}$.

Inputs are often evaluated at their hystorical value (e.g. the price at which the input was bought) so that they are not influenced by current prices.

Hence, presumptive output for firm i belonging to the population of active firms in tax year t is calculated as $y_{it} = \mathbf{b}'_t \mathbf{x}_{it}$. Notice that reported input ($\mathbf{x}_{it}$) and output (y_{it}) of firm i can differ from their true values, which we denote by $\widetilde{\mathbf{x}}_{it}$ and $\widetilde{y}_{it}$, respectively.

We write perceived probability to be audited as

$$p_{it}\left(y_{it}, \mathbf{x}_{it}\right) \tag{2}$$

where $(y_{it}, \mathbf{x}_{it})$ is the signal reported by the firm to the TA. We assume that the signal is increasing in y_{it} and decreasing in $\mathbf{x}_{it}$, i.e. that, *ceteris paribus*, the probability to be audited decreases as reported output increases and increases as reported input increases. Note that we index the probability function thus allowing for subjective probabilities.

The relationship between $y_{it}, \mathbf{b}_t$ and $\mathbf{x}_{it}$ defines the inconsistency status of the firm: a firm is said to be not consistent (*incongrua*) when $y_{it} < \mathbf{b}'_t \mathbf{x}_{it}$ and consistent (*congrua*) when $y_{it} \geq \mathbf{b}'_t \mathbf{x}_{it}$, so that an inconsistency dummy D_{it} for firm i in period t is defined as follows

$$D_{it} = \begin{cases} 1 \text{ if } y_{it} < \mathbf{b}'_t \mathbf{x}_{it} \\ 0 \text{ if } y_{it} \geq \mathbf{b}'_t \mathbf{x}_{it} \end{cases} \tag{3}$$

The TA uses Sds as a method to select taxpayers to be audited. In the event a taxpayer is audited, the audit concerns the difference between reported and true output, if any[2].

[2] In this paper we interpret p and the inconsistency status differently from Santoro and Fiorio (2011) to account for some recent changes in the institutional and legal context. Santoro and Fiorio (2011) assumed that $p = 0$ when $D_{07} = 0$ and that $1 > p > 0$ when $D_{07} = 1$,with p decreasing in the ratio $y/\mathbf{b}'\mathbf{x}$ for values belonging to the interval $(0, 1')$.This assumption was motivated by the use of Sds to determine jointly the audit probability *and*

To complete the description, a fundamental piece of information concerns the timing of the game. For reasons discussed in Santoro (2008), Sds has been designed so that $\mathbf{b}_t$ is fully known when $\mathbf{x}_{it}$ is reported by firm i. In practice, firms are asked to report input (and output) values using a software (known as Ge.ri.co) which contains full information on the value of each element of $\mathbf{b}_t$. By using this software, every firm i, or more likely its tax consultant, can try different values of $(\mathbf{x}_i, y_i)$ to minimize expected tax payments. In particular, since usually $b^j > 0$ $(\forall j = 1, 2, ..., J)$ the more common manipulation is the underreporting of $\mathbf{x}_i$ with respect to its true value $\widetilde{\mathbf{x}}_i$ to decrease due tax holding audit probability to a minimum.

The intent to detect and disincentive input underreporting is the primary objective of the letter campaign which we focus on in this paper. Before turning to its analysis, let us define the vector of input manipulation input by firm i for tax period t as $\mathbf{m}_{it} = (\widetilde{x}_{jit} - x_{jit})$. Thus, the manipulation dummy variable M_{it} can be defined as

$$M_{it} = \begin{cases} 1 \text{ if there exists at least one } j \in \{1, 2, .., J\} \text{ s.t. } m_{jit} > 0 \\ 0 \text{ otherwise} \end{cases}$$

Finally, we define a letter dummy variable L_{it} that takes value equal 1 if a letter is sent to firm i for tax year t and takes value equal 0 otherwise. These two variables are logically related, since the letter is sent to those firms which, according to the information available to the TA, are alleged to have manipulated inputs in 2007. However, we will briefly discuss the possibility that the letter is sent by mistake, i.e. that $L_{it} = 1$ and $M_{it} = 0$.

4 The letter campain

At the beginning of 2009, i.e some months before issuing their tax reports referring to tax year 2008, more than 100,000 businesses (firms and professionals) received a letter from the TA informing them that:

the amount due by the firm in case of an audit. In this perspective, the firm which reports a ratio $y/\mathbf{b}'\mathbf{x} < 1$ could be audited and asked to pay (only) the difference between y and $\mathbf{b}'x$. Actually, in such a context, the term 'audit' was itself unappropriate, since the Tax Agency was not assumed to be searching for the true value of output, but, rather, to implement its presumptive value. This presumptive value was thus given a legal strength so that Sds turn to a method of 'normal' or 'minimum' taxation rather than a method to calculate the audit probability function. Accordingly, no taxpayer reporting $y_i \geq \mathbf{b}'x_i$ could be 'audited' and this explains why Santoro and Fiorio (2011) assume that $p = 0$ when $D_{07} = 0$.In recent years, the legitimity of this use of Sds has been questioned by tax lawyers and, in some cases, tax judges and courts'rulings have denied the taxpayer's obligation to pay the difference (if any) between y and $\mathbf{b}'x$.

- **a)** some input reports (x_{ji}) they made for tax year 2007 were deemed to be "anomalous";

  b) if this anomaly or a similar one was to be repeated for tax year 2008, i would certainly be included in a list of taxpayers to be audited.

Considering the structure of Sds described in Section 3, the aim of the letter is quite clear: the business is informed that his practice of underreporting inputs to decrease the probability to be audited has been detected. There are two main differences between the letter campaign we study here and examples reported in Section 2. First, the Italian campaign has a different objective with respect to those examined in the literature, since it is not directly purported to elicit higher output reports. The business should primary perceive an increase in the audit probability if it persists to manipulate inputs. So, the first (expected) reaction is that it raises its report of input, rather than output. However, this behaviour will automatically decrease its signal and thus the overall probability to be audited will increase altogether. Thus, a business could finally decide to increase *both* its output and input reports. We may say that this is the real objective of the TA, i.e. to push firms to report higher output and to pay higher taxes via an increased perceived probability to be audited.

However, and this is the second difficulty, while one can speculate that input reporting does not depend on economy-wide considerations, this clearly does not apply to output reports. More specifically, if one has to compare output reports in tax year 2008 to output reports in tax year 2007 by firms which received the letter a counterfactual is needed. Ideally, one should compare the change in output reported in tax year 2008, with respect to 2007, by a random sample of firms which received the letter to a random sample of firms which did not receive the letter. Unfortunately, the Italian campaign was not designed as a field experiment and randomization was not adopted: letters were sent to all firms which allegedly manipulated input values. This clearly poses some methodological issues that we deal with in Section 6.

In this Section we try to explain why and under which premises the letter can be expected to elicit an higher output report.

Taking into account the effect of the manimulation activity, we can rewrite the subjective probability to be audited in tax year 2008 for a manipulating firm which received the letter as

$$p_{08} = p(y_{08}, x_{08} \mid M_{07} = 1, L_{08} = 1) \tag{4}$$

with p_{08} increasing in m_{08}. Note that we are assuming separability of the audit probability function in 2008 in its two arguments, i.e. the signal and the amount of input manipulation. The probability to be audited in tax year 2008 for a firm which did not receive the letter can be written as

$$p_{08} = p(y_{08}, x_{08} \mid M_{07} = 0, L_{08} = 0) \tag{5}$$

If we set aside for a moment the possibility that the letter was received by mistake, i.e. that it was sent to a firm which in fact was not not manipulating inputs ($M_{07} = 0, L_{08} = 1$),and if we assume that $\mathbf{b_{07}} = \mathbf{b}_{08}$,[3] then the impact of the letter is to decrease the probability to be audited if the firm does not reduce input manipulation. Thus, a firm can be induced to decrease input manipulation, i.e. to reduce m_{08} with respect to m_{07}. However, by doing so, *if output report is unchanged*, the signal will be reduced and, therefore, $p_{08} > p_{07}$.Thus, the firm may be induced to increase output report, i.e. to report $y_{08} > y_{07}$ as a reaction to the letter.

Clearly, a change in output report can be driven by many other factors, such as a change in the market and demand conditions. To capture the difference between the impact of the letter and of all other exogenous possible factors we compare the change in reported output by firms which received the letter to that by firms which did not receive the letter. If the letter has an impact on output reporting we expect to find a significant and positive difference between these two changes. However, this result can be altered by the presence of many letters sent by mistake, i.e. if there were many instances where $M_{07} = 0$ and $L_{08} = 1$.In such a case there are at least two possible reactions by the firm. Considering that the audit is on the difference between reported and true output, a firm, even if it has not manipulated inputs, could choose to report $x > \widetilde{x}$ just to decrease the probability of being audited (provided that the impact on m is higher than the opposite impact on s). Alternatively, a firm which has not manipulated inputs could feel to be in a safe position and to ignore the letter altogether. In this second case, an absence of any significant impact of the letter on output reports is a logical consequence.

5 Data description

In this paper we use a data set produced by the Italian TA for this project with the aim of estimating the effectiveness of the letter campaign on declared

[3] Actually $\mathbf{b}'_{08}$ is estimated using data reported by firms which were reliable in 2005, which are usually different from firms which were reliable in 2004. However we ignore this difference here since we do not have data on such a change.

profit and revenues. The data set provided contains a sample of 49,138 treated firms and a sample of 89,240 controls.

The sample of treated firms is randomly extracted from an initial sample of over 112,000 corporated, incorporated firms and professionals who were suspected to have manipulated inputs in year 2007, according to some indicators developed by the TA and not fully available to taxpayers nor to us. For this sample we have information on:

a) a set of characteristics regarding location area (in five major areas, North-West, North-East, Center, South, Islands), the business sector, the legal form (whether self-employed professional, firm using simplified or standard accounting methods);

b) data on costs of inputs, services, costs for purchased services, intermediate goods, inventories, labour services, the number of dependent workers distinguished into full time permanent, full time temporary workers, family and non-family collaborators, as well as declared profit and revenues;

c) the level of reported and presumptive output and the type of anomaly recorded into 19 categories, provided by the TA and pointed out in the letter campaign to addressed taxpayers.

The sample of controls is randomly extracted from an initial sample of over 2,2 millions of firms which were not suspected to have manipulated inputs.

Our identification strategy regarding the effect of the letter on output and profits is based on matching treated firms with untreated ones based on a set of characteristics observed prior to treatment, as data do not come from a field experiment and all firms who – according to some TA indicators – allegedly manipulated inputs in tax year 2007 received the letter requiring action. For this aim the TA provided us with full information about all treated and control firms in our sample regarding tax year 2006, i.e. the tax year before to treatment, when no campaign was implemented nor announced, yet.

Finally, we were also given information as for tax year 2008, i.e. after treatment, which we use extensively to assess the causal effect of the treatment.

Table 1 reports some descriptive statistics for the treated and control sample separately, in 2006, the year just before treatment. Treated units are more likely to be located in the South and the Islands, are more likely to use standard accounting methods. Treated firms are also more likely to be operating in the construction and in the trade sectors. As for inventories, it clearly emerges that treated firms have much higher average levels of both beginning and ending inventories, with higher revenues but lower profits, while differences in the size of the firms' workforce seem negligible.

6 The matching method: coarsened exact matching

As described above, the letter campain was not properly designed as a field experiment as it main aims was to induce people to reduce input manipulation and eventually to increase tax revenues, rather than finding the effect of a deterrence policy. Hence we have to revert to some matching method, which could be described as a nonparametric method to control for the confounding influence of pretreatment control variables in observational data. The main aim of matching is to prune observations from the data so that the remaining data have better balance between the treated and control groups. In other words, a better balance can be described as the fact that the empirical distributions of the covariates ($\mathbf{X}$) in the treated and control groups are more similar. In case of exactly balanced data, controlling further for $\mathbf{X}$ is not necessary as it is not correlated to the treatment variable, and a a simple difference in means on the matched data can provide an estimate of the causal effect. Differently, approximately balanced data require controlling for $\mathbf{X}$ with a model.

As extensively discussed in Iacus, King, and Porro (2011a), central dilemma means that model dependence and statistical bias are usuallymuch bigger problems than large variances and most matching methods seem designed for the opposite problem. In fact, they guarantee the matched sample size ex ante (thus fixing most aspects of the variance) and produce some level of reduction in imbalance between the treated and control groups (hence reducing bias and model dependence) only as a consequence and only sometimes. As they put it " [...] the less important criterion is guaranteed by the procedure, and any success at achieving the most important criterion is uncertain and must be checked ex post. Because the methods are not designed to achieve the goal set out for them, numerous applications of matching methods fail the check and so need to be repeatedly tweaked and rerun." (p. 2).

To avoid these and other problems with most existing matching methods Iacus, King, and Porro (2011b) , introduce a new generalized class of matching methods known as "Monotonic Imbalance Bounding" of which we use the "Coarsened Exact Matching" (CEM). CEM works in sample and re-

	Sample of treated						Sample of controls				
Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
						Area of location (omitted category: North-West)					
North-East	49138	0.169	0.375	0	1	North-East	89240	0.242	0.428	0	1
Center	49138	0.206	0.404	0	1	Center	89240	0.208	0.406	0	1
South	49138	0.257	0.437	0	1	South	89240	0.169	0.375	0	1
Islands	49138	0.118	0.322	0	1	Islands	89240	0.077	0.266	0	1
						Additional productive locations (omitted category: no additional locations)					
One add. location	49138	0.757	0.429	0	1	One add. location	89240	0.830	0.376	0	1
Two or more add. locat.	49138	0.068	0.251	0	1	Two or more add. locat.	89240	0.060	0.238	0	1
						Accounting methods (omitted category: simplified accounting)					
Standard accounting	49138	0.449	0.497	0	1	Standard accounting	89240	0.401	0.490	0	1
Not-for-profit firms	49138	0.001	0.023	0	1	Not-for-profit firms	89240	0.000	0.021	0	1
						Sector of operation (omitted category: Industry)					
Agriculture	49138	0.001	0.037	0	1	Agriculture	89240	0.002	0.043	0	1
Construction	49138	0.171	0.377	0	1	Construction	89240	0.158	0.365	0	1
Wholesale trade	49138	0.195	0.396	0	1	Wholesale trade	89240	0.158	0.364	0	1
Retail trade	49138	0.274	0.446	0	1	Retail trade	89240	0.207	0.405	0	1
Transportation	49138	0.007	0.081	0	1	Transportation	89240	0.048	0.213	0	1
Hotels and restaurants	49138	0.078	0.269	0	1	Hotels and restaurants	89240	0.078	0.268	0	1
Computer activities	49138	0.026	0.161	0	1	Computer activities	89240	0.022	0.148	0	1
Financial intermediation	49138	0.023	0.150	0	1	Financial intermediation	89240	0.011	0.105	0	1
Real estate activities	49138	0.049	0.215	0	1	Real estate activities	89240	0.057	0.232	0	1
Other professionals	49138	0.015	0.121	0	1	Other professionals	89240	0.022	0.145	0	1
Other services	49138	0.041	0.199	0	1	Other services	89240	0.097	0.295	0	1
Health services	49138	0.002	0.042	0	1	Health services	89240	0.004	0.067	0	1
						Inventories and financial variables (in 000 euro)					
Beg. inv., work in prog.	49138	29.485	379.157	0	61915.690	Beg. inv., work in prog.	89240	4.929	101.976	0	6580.909
Beg. inv., finished goods	49138	8.455	128.009	0	6566.878	Beg. inv., finished goods	89240	0.770	42.734	0	5907.026
Ending inventory	49138	142.943	445.340	0	14882.080	Ending inventory	89240	56.414	277.451	0	37880.490
Capital goods value	49138	82.993	256.773	0	7331.896	Capital goods value	89240	90.250	250.797	0	6668.831
Costs, row materials	49138	165.123	394.369	0	4748.488	Costs, row materials	89240	126.638	348.288	0	6373.854
Costs, finished goods	49138	11.114	37.514	0	2197.878	Costs, finished goods	89240	12.122	42.025	0	3811.614
Costs, services	49138	60.755	198.742	0	6303.979	Costs, services	89240	53.457	166.757	0	4909.338
Costs, labour	49138	45.379	128.791	0	3670.223	Costs, labour	89240	41.614	119.211	0	3956.965
Depreciation	49138	9.106	28.147	0	1434.158	Depreciation	89240	9.642	29.053	0	2184.034
Provisions	49138	0.244	5.029	0	636.000	Provisions	89240	0.235	6.125	0	1400.404
Financial result	49138	-3.919	45.828	-2380.678	3659.134	Financial result	89240	-2.486	32.840	-3318.388	2379.357
Revenues	49138	319.515	620.665	0	7259.844	Revenues	89240	282.177	563.945	0	23016.400
Profits	49138	29.065	108.978	-5764.723	10590.750	Profits	89240	32.970	103.533	-4183.722	21219.220
Marginal tax rate	49138	6.514	11.526	0	27.500	Marginal tax rate	89240	4.874	10.270	0	27.500
No. of FT empl.	49138	1.555	3.885	0	89.038	No. of FT empl.	89240	1.490	3.715	0	128.955
No. of temp workers	49138	0.138	0.972	0	77.000	No. of temp workers	89240	0.133	0.901	0	91.000
No. of family members	49138	0.100	0.402	0	33.000	No. of family members	89240	0.141	0.439	0	19.000

Source: Our calculations on TA data.

Table 1: Descriptive statistics for treated and control groups before the treatment, i.e. in year 2006.

quires no assumptions about the data generation process (beyond the usual ignorability assumptions). More importantly, CEM inverts guarantees that the imbalance between the matched treated and control groups will not be larger than the ex ante user choice. CEM is matching method that allows researchers to choose the maximum imbalance between the treated and control groups ex ante, rather than assed through the usual process of ex post checking and repeatedly reestimating. CEM bounds through ex ante user choice both the degree of model dependence and the average treatment effect estimation error, eliminating the need for a separate procedure to restrict data to common empirical support. It also meets the congruence principle, is robust to measurement error and is fast computationally even with very large data sets.

CEM as all matching methods can be described as a way to preprocess a data set so that estimation of the sample average treatment on the treated (ATT), based on the matched data set, will be less a function of apparently small and indefensible modeling decisions, than when based on the original full data set. Matching involves pruning observations that have no close matches on pre-treatment covariates in both the treated and control groups, resulting in less model-dependence, bias, and inefficiency (King and Zeng 2006, Ho, Imai, King, and Stuart 2007, Iacus, King, and Porro 2011b).

In this paper we apply CEM requiring the assumption of ignorability (a.k.a. "no omitted variable bias" or "no confounding"). Our specific statistical goal is to estimate the causal effect of the letter campain on average, or the sample ATT. Let Y_i be the dependent variable for unit i, which in our case is the log difference or the log ratio of output or of profits between year 2008 and year 2007. Let T_i be a dichotmous treatment variable taking value 1 for treated and 0 for control units, and X_i be a vector of pre-treatment control variables, which includes the variables described in Section 5. The average treatment effect for treated units is then the difference between two potential outcomes: $TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)$, where $Y_i(T_i = 1) = Y_i$ is always obseved and $Y_i(T_i = 0)$, the value that Y_i would have taken on if it were the case that $T_i = 0$, is always unobserved. Then $Y_i(T_i = 0)$ we estimate with Y_j from matched controls (i.e., among units for which $X_i \approx X_j$) directly, $\widehat{Y}_i(T_i = 0) = Y_j(T_j = 0)$, avodingin using a discretional model, $\widehat{Y}_i(T_i = 0) = \widehat{g}(X_j)$. Then the ATT will then computed as a simple average:

$$ATT = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} TE_i \tag{6}$$

Interestingly, Iacus, King, and Porro (2011b) also introduced a simple and comprehensive multivariate imbalance measure of the actual degree of imbalance achieved in the matched sample, which may be lower than the chosen maximum. The measure is based on the L^1 difference between the multidimensional histogram of all pretreatment covariates in the treated group and that in the control group, which is used to evaluate improvements in matching imbalance with different matching methods. This measure is computed by cross-tabulating the discretized variables as $X_1 \times \cdots \times X_k$ for the treated and control groups separately, and recording the k-dimensional relative frequencies for the treated $f_{\ell_1 \ldots \ell_k}$ and control $g_{\ell_1 \ldots \ell_k}$ units. Hence, the measure of imbalance is computed as the absolute difference over all the cell values:

$$L^1(f,g) = \frac{1}{2} \sum_{\ell_1 \ldots \ell_k} |f_{\ell_1 \ldots \ell_k} - g_{\ell_1 \ldots \ell_k}| \tag{7}$$

and where the summation is over all cells of the multivariate histogram, but is feasible to compute because it contains at most n nonzero terms. The L^1 measure1 varies in $[0,1]$. Perfect (up to discretization) global balance results in $L^1 = 0$, and $L^1 = 1$ indicates complete separation of the multimensional histograms. Any value in the interval $(0,1)$ indicates the amount of difference between k-dimensional frequencies of the two groups.

7 Empirical results

Table 2 shows the (7) measure of imbalance for four different coarsening produced on our data set. In particular, column (1) presents the automated coarsening which provides a measure of imbalance which we use as reference to assess the effectiveness of following coarsening. It shows a level of imbalance of 0.91,which is relatively close to the maximum of 1. This also allows us to have a relatively small level of pruning, with few unmatched treated observations. Column (2) presents the thinnest coarsening, hence providing the highest level of pruning, with about two thirds of the treated sample resulting unmatched. Columns (3) and (4) present intermediate levels of coarsening. Notice that by deciding the level of coarsening we are able to largely reduce the overall measure of imbalance.

Hence, by using these different CEM procedure we estimate the ATT of the letter campain, where the dependent variable is defined as the log of the ratio of output (or profit) in year 2008 over that of year 2007. Table 3 shows that the letter had a statistically signifincant impact on output reported. The log of the ratio between output reported in 2008 and output reported in 2007 is significantly higher for treated firms with respect to controls. Results show that, on average, the percentage variation in output reported by treated firms is higher from a minimum of 1,014 times to a maximum of 1,055 times than the same percentage variation as reported by firms which did not receive the letter. Results are always and highly statistically significant and are basically unaltered when profit rather than output reports are considered.

Overall, these results seem to suggest that firms have reacted to the letter not simply by reducing manipulation, i.e. increasing reported inputs, but also, and more importantly, by increasing reported output and profits. Since the latter are taxable income, this means that the letter campaign has probably produced a net increase in taxes paid.

Coarsened exact matching				
	(1)	(2)	(3)	(4)
Imbalance	0.910	0.475	0.602	0.623
Observations	138,378	59,176	69,425	66,169

Source: Our calculations on TA data.

Table 2: Multivariate imbalance measure after different coarsened exact matching procedures.

8 Discussion and concluding remarks

OUTPUT

	(1)	(2)	(3)	(4)
Treated	0.014*	0.056***	0.050***	0.054***
	[0.007]	[0.011]	[0.010]	[0.011]
Constant	-0.040***	-0.059***	-0.057***	-0.055***
	[0.004]	[0.007]	[0.006]	[0.007]
Observations	138,378	59,176	69,425	66,169

PROFITS

	(1)	(2)	(3)	(4)
Treated	0.036***	0.057***	0.050***	0.054***
	[0.009]	[0.013]	[0.012]	[0.013]
Constant	-0.058***	-0.083***	-0.077***	-0.074***
	[0.005]	[0.008]	[0.008]	[0.008]
Observations	138,378	59,176	69,425	66,169

Notes: Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1
Source: Our calculations on TA data.

Table 3: Causal effect estimation of the letter campain on the log of the ratio of revenues and profits, between 2008 and 2007.

References

Blumenthal, M., C. Christian, and J. Slemrod (2001). Do normative appeals affect tax compliance? evidence from a controlled experiment in minnesota. *National Tax Journal 54*(1), 125 – 138.

Ho, D., K. Imai, G. King, and E. Stuart (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis 15*, 199 – 236.

Iacus, S. M., G. King, and G. Porro (2011a). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*.

Iacus, S. M., G. King, and G. Porro (2011b, 2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association 106*, 345–361.

King, G. and L. Zeng (2006). The dangers of extreme counterfactuals. *Political Analysis 14*(2), 131 – 159.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*(3), 651–692.

Macho-Stadler, I. and J. D. Pérez-Castrillo (2002). Auditing with signals. *Economica 69*(273), pp. 1–20.

Santoro, A. (2008, July). Taxpayers' choices under studi di settore:what do we know and how we can interpret it? *Giornale degli Economisti 67*(2), 161–184.

Santoro, A. and C. V. Fiorio (2011, January). Taxpayer behavior when audit rules are known: Evidence from italy. *Public Finance Review 39*(1), 103–123.

Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in minnesota. *Journal of Public Economics 79*(3), 455–483.